18.405J/6.841J Advanced Complexity Theory	February 29, 2000

Lecture 8

Lecturer: Daniel A. Spielman — *Scribe: Mohammad Mahdian*

In this lecture, we will prove the theorem of Valiant and Vazirani, and introduce the complexity of counting.

8.1 The Valiant-Vazirani Theorem

The Valiant-Vazirani Theorem [1] says that if we can solve SAT instances with unique solutions, then we can solve SAT in general (using randomization). The following is the exact statement of their theorem.

Theorem 1 *There is a randomized polynomial time Turing machine M that on input $\phi \in$ 3-CNF with n variables, M outputs $n+2$ 3-CNFs $\phi_1, \phi_2, \ldots, \phi_{n+2}$ such that*

(i) if $\phi \in \mathsf{SAT}$, then $\Pr[\exists i \text{ such that } \phi_i \text{ has exactly one satisfying assignment}] \geq \frac{1}{8}$.

(ii) if $\phi \notin \mathsf{SAT}$, then $\forall i\ \phi_i \notin \mathsf{SAT}$.

The idea of the proof is to add some additional constraints randomly to the formula ϕ to get several formulas with fewer satisfying assignments, so that if ϕ is satisfiable, with high probability at least one of those formulas have exactly one satisfying assignment.

We need the concept of pairwise independent hash functions that was introduced in the problem set 1 (problem 5).

Definition 2 *A family $\mathcal{H}$ of functions from $\{0,1\}^n$ to $\{0,1\}^k$ is* pairwise independent *if*

- *for every $x \in \{0,1\}^n$ and $a \in \{0,1\}^k$,*

$$\Pr[h(x) = a] = \frac{1}{2^k},$$

- *and for every $x, y \in \{0,1\}^n$, $x \neq y$, and $a, b \in \{0,1\}^k$,*

$$\Pr[h(x) = a \mid h(y) = b] = \frac{1}{2^k}.$$

From Problem Set 1, Problem 5, we know that there are families of pairwise independent hash functions that are easy to compute, i.e., it is possible to construct a circuit computing a random element of the family in randomized polynomial time. In particular, the family

$$\mathcal{H}_{n,k} = \{h_{M,s} : M \in GF(2)^{n \times k}, s \in GF(2)^k, h_{M,s}(x) = xM + s\} \tag{8.1}$$

is pairwise independent.

Now, assume that we know that the formula ϕ has between 2^{k-2} and 2^{k-1} solutions. We can build a formula ϕ' that is satisfied by $\overline{x}$ if and only if

- ϕ is satisfied by $\overline{x}$, and
- $h(\overline{x}) = \overline{0}$, where h is chosen from a pairwise independent family $\mathcal{H}$ at random.

We hope that ϕ' has just one satisfying assignment. The following lemma justifies this fact.

Lemma 3 *Let S be a subset of $\{0,1\}^n$ such that $2^{k-2} \leq |S| \leq 2^{k-1}$, and $\mathcal{H}$ be a pairwise independent family of hash functions from $\{0,1\}^n$ to $\{0,1\}^k$. Then*

$$\Pr_{h \in \mathcal{H}}[|S \cap h^{-1}(\overline{0})| = 1] \geq \frac{1}{8}.$$

Proof: We have:

$$\Pr_{h \in \mathcal{H}}[|S \cap h^{-1}(\overline{0})| = 1] = \Pr_{h \in \mathcal{H}}[\exists x \in S \text{ such that } (h(x) = \overline{0} \wedge \forall y \in S - \{x\}, h(y) \neq \overline{0})]$$

Since the above events are mutually exclusive, the above probability is equal to:

$$\begin{aligned}
\Pr_{h \in \mathcal{H}}[|S \cap h^{-1}(\overline{0})| = 1] &= \sum_{x \in S} \Pr_{h \in \mathcal{H}}[h(x) = \overline{0} \wedge \forall y \in S - \{x\}, h(y) \neq \overline{0}] \\
&= \sum_{x \in S} \Pr_{h \in \mathcal{H}}[h(x) = \overline{0}] \Pr_{h \in \mathcal{H}}[\forall y \in S - \{x\}, h(y) \neq \overline{0} \mid h(x) = \overline{0}] \\
&= \sum_{x \in S} \frac{1}{2^k}(1 - \Pr_{h \in \mathcal{H}}[\exists y \in S - \{x\}, h(y) = \overline{0} \mid h(x) = \overline{0}]) \\
&\geq \sum_{x \in S} \frac{1}{2^k}(1 - \sum_{y \in S - \{x\}} \Pr_{h \in \mathcal{H}}[h(y) = \overline{0} \mid h(x) = \overline{0}]) \\
&\geq \sum_{x \in S} \frac{1}{2^k}(1 - \frac{|S|}{2^k}) \\
&= \frac{|S|}{2^k}(1 - \frac{|S|}{2^k}) \geq \frac{1}{4} \times \frac{1}{2} = \frac{1}{8}.
\end{aligned}$$

■

Lemma 4 *Let $\mathcal{H}_{n,k}$ be the family defined by Equation 8.1. For any $h \in \mathcal{H}_{n,k}$ we can construct in polynomial time a 3-CNF ϕ_h such that*

(i) if $h(\overline{x}) = 0$, then there is a unique $\overline{y}$ such that $\phi_h(\overline{x}, \overline{y}) = 0$.

(ii) if $h(\overline{x}) \neq 0$, then for every $\overline{y}$, $\phi_h(\overline{x}, \overline{y}) \neq 0$.

We don't prove the above lemma, but the proof is easy (It is sufficient to construct a circuit that computes the function h, and then turn the circuit into a formula).

Now, we are ready to prove the theorem of Valiant and Vazirani.

Proof of Theorem 1: For $k = 1, \ldots, n$, choose a random hash function $h_k : \{0,1\}^n \mapsto \{0,1\}^k$ from $\mathcal{H}_{n,k}$ and set $\phi_k(\overline{x}) = \phi(\overline{x}) \wedge \phi_{h_k}(\overline{x}, \overline{y})$. If the number of satisfying assignments to ϕ is between 2^{k-2} and 2^{k-1}, then by Lemma 3 and Lemma 4, ϕ_k has just one satisfying assignment with probability at least $\frac{1}{8}$. ■

We can formulate this result in terms of the complexity classes. We need the following definition.

Definition 5 *For a predicate Q, the predicate USAT_Q is defined as follows:*

$$\mathsf{USAT}_Q(\phi) = \begin{cases} 0 & \text{if } \phi \notin \mathsf{SAT} \\ 1 & \text{if } \phi \text{ has a unique satisfying assignment} \\ Q(\phi) & \text{otherwise} \end{cases}$$

The following theorem is a simple corollary of Theorem 1.

Theorem 6 *If there is a Q such that $\mathsf{USAT}_Q \in \mathbf{RP}$, then $\mathbf{NP} \subseteq \mathbf{RP}$.*

8.2 The complexity of counting solutions

In 1979, Valiant asked how hard is it to count the number of solutions. The above result of Valiant and Vazirani helped to prove Toda's theorem which says that counting the number of solutions is at least as difficult as the polynomial hierarchy.

Definition 7 *$\#\mathbf{P}$ is the class of functions $f : \{0,1\}^n \mapsto N$ for which there is a nondeterministic polynomial time Turing machine M such that for every word w, $M(w)$ has exactly $f(w)$ accepting paths.*

The language $\#\mathsf{SAT}$ that is defined below is the complete problem for $\#\mathbf{P}$ (The proof of this fact, together with the precise definition of $\#\mathbf{P}$-completeness will be presented in the next lecture).

Definition 8 *For every formula ϕ, $\#\mathsf{SAT}(\phi)$ is the number of satisfying assignments of ϕ.*

Since SAT is an $\mathbf{NP}$-complete problem, it is natural to expect $\#\mathsf{SAT}$ to be $\#\mathbf{P}$-complete; but Valiant proved that it is also possible to get $\#\mathbf{P}$-complete problems out of problems that are polynomially solvable:

Theorem 9 *Let $\#\mathsf{Matching}$ be the function that computes the number of perfect matchings of the given graph. Then $\#\mathsf{Matching}$ is a $\#\mathbf{P}$-complete problem.*

Since $\#\mathbf{P}$ is a class of functions, we cannot compare it with the other complexity classes, which are classes of languages. Therefore, we consider the class $\mathbf{P}^{\#\mathbf{P}}$. It is clear that $\mathbf{NP}$ and $\mathbf{co\text{-}NP}$ are both contained in $\mathbf{P}^{\#\mathbf{P}}$. In 1989, Toda proved the following.

Theorem 10 $\mathbf{PH} \subseteq \mathbf{P}^{\#\mathbf{P}}$.

References

[1] L. G. Valiant, and V. V. Vazirani, **NP** *is as easy as detecting unique solutions*, Theoretical Computer Science **47** (1986), no. 1, 85–93.